



# A rapid, whole genome sequencing and analysis system supporting molecular epidemiology

Michael FitzGerald

Broad Institute

SFAF 2012

# Collaborators

- **BI** : Yonatan Grad
- **HSPH**: Bill Hanage, Marc Lipsitch
- **Statens Institute, Copenhagen, Denmark**: Jakob Frimodt-Møller, Andreas M. Petersen, Carsten Struve, Karen A. Krogfelt
- **Institut Pasteur, Paris, France**: François-Xavier Weill
- **Laboratoire de Microbiologie, Hôpital Robert Debré, Paris, France**: Edouard Bingen
- **Broad Institute**: Deb Hung, Michael Feldgarden, Harindra Arachchi, Gustavo Cerqueira, Paul Godfrey, Brian Haas, Cheryl Murphy, Carsten Russ, **Sean Sykes**, **Bruce Walker**, Jennifer Wortman, **Sarah Young**, Qiandong Zeng, Amr Abouelleil, James Bochicchio, Sara Chauvin, Timothy DeSmet, Sharvari Gujja, Caryn McCowan, **Anna Montmayeur**, Scott Steelman, Eric Lander, Chad Nusbaum, Bruce Birren

# Repurposing a genome factory



96 libraries per run  
10+ weeks turnaround



10's of libraries  
Days turnaround

# Content

- E. coli O104:H4 project
- Hijacking a high throughput genome sequencing system
- Fire drill system
- Acinetobacter project

June 2, 2011

# E. Coli Strain Was Previously Unknown, Official Says

By ALAN COWELL and JAMES KANTER

BERLIN — The strain of E. coli setting off alarms throughout Europe is a previously unknown “super-toxic” variant, international health officials said Thursday.

The New York Times

June 3, 2011



Robert Solsona/Associated Press

Farmers dumped produce at the German consulate in Valencia to protest claims that the E. coli outbreak started in Spain.



June 10, 2011

# Germany Says Bean Sprouts Are Likely E. Coli Source

By ALAN COWELL

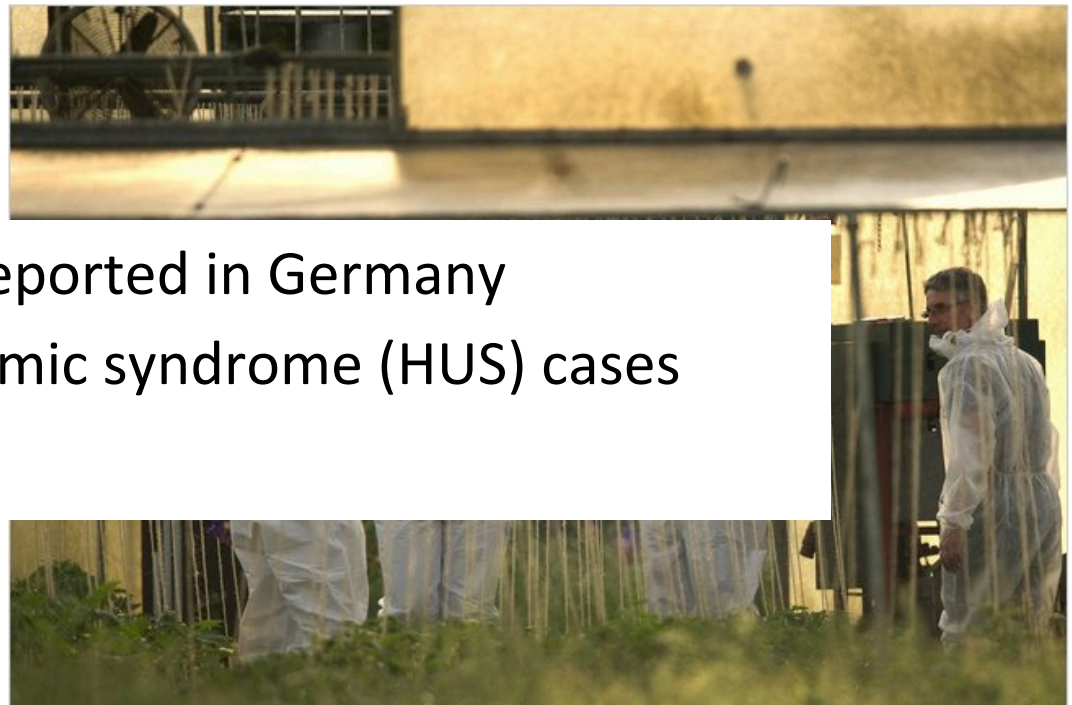
BERLIN — After days of confusion, German authorities said on Friday that they had concluded that contaminated sprouts from an organic farm in the country's north were the most likely cause of one of the world's worst outbreaks of E. coli.

The New York Times

June 11, 2011



- 3816 total cases reported in Germany
- 845 hemolytic uremic syndrome (HUS) cases
- 54 deaths



Odd Andersen/Agence France-Presse - Getty Images

Investigators on Monday examined a farm in Bienenbüttel, Germany, that is believed to be the source of contaminated sprouts.

# Outbreak isolates are indistinguishable by conventional molecular epidemiology

## Virulence factors MLST

Product		Presence in <i>E. coli</i> O104:H4*	Presence in HUSEC041	Presence in EAEC 55989†
STEC				
stx <sub>1</sub>	Shiga toxin 1	0/80	No	No
stx <sub>2</sub>	Shiga toxin 2	80/80	Yes	No
EHEC-hlyA	EHEC haemolysin	0/80	No	No
cdt (I-V)	Cytolethal distending toxin	0/80	No	No
subAB	Subtilase cytotoxin	0/80	No	No
espP	Serine protease EspP	0/80	No	No
eae	Intimin	0/80	No	No
iha	Iha (IrgA homologue adhesin)	80/80	Yes	Yes
lpfA <sub>O26</sub>	Structural subunit of long polar fimbriae (LPF) of STEC O26	80/80	Yes	Yes
lpfA <sub>O113</sub>	Structural subunit of LPF of STEC O113	80/80	Yes	Yes
lpfA <sub>O157-O154</sub>	Structural subunit of LPF of STEC O157:H7 (encoded on O island 141)	0/80	No	No
lpfA <sub>O157-O154</sub>	Structural subunit of LPF of STEC O157:H7 (encoded on O island 154)	0/80	No	No
saa	Saa (STEC autoagglutinating adhesin)	0/80	No	No
sfpA	Structural subunit of Sfp fimbriae	0/80	No	No
ter cluster	Tellurite resistance	80/80	Yes	No
irp2	Component of iron uptake system on HPI	80/80	Yes	No
fyuA	Component of iron uptake system on HPI	80/80	Yes	No
EAEC				
aatA	EAEC virulence plasmid (pAA)‡	80/80	Yes	Yes
aggA	Pilin subunit of aggregative adherence fimbriae I (AAF/I)	80/80	No	No
agg3A	Pilin subunit of aggregative adherence fimbriae III (AAF/III)	0/80	Yes	Yes
aggR	Transcriptional regulator AggR	80/80	Yes	Yes
aap	Dispersin	80/80	Yes	Yes
set1	<i>Shigella</i> enterotoxin 1	80/80	Yes	Yes
pic	Pic (protein involved in intestinal colonisation)	80/80	Yes	Yes
astA	EAEC heat-stable enterotoxin 1 (EAST1)	0/80	Yes	Yes
EPEC				
bfpA	Bundle-forming pili	0/80	No	No
ETEC				
elt	Heat-labile enterotoxin (LT)	0/80	No	No
estIa	Heat-stable enterotoxin (STIa)	0/80	No	No
estIb	Heat-stable enterotoxin (STIb)	0/80	No	No
EIEC				
ial	Invasive plasmid (pInV) <sup>§</sup>	0/80	No	No

STEC=Shiga toxin-producing. EPEC=enteropathogenic. EAEC=enteroaggregative. ETEC=enterotoxigenic. EIEC=enteroinvasive *E. coli*. HPI=high pathogenicity island. \*Number of strains positive for the locus/number tested. †Based on in silico analysis of the sequence of EAEC strain 55989 (GenBank accession number NC\_011748). ‡Tested by PCR targeting the region of the EAEC virulence plasmid used as an EAEC diagnostic probe (pCVD432)<sup>15</sup> for *aatA*.<sup>20</sup> §Tested by PCR targeting the region of pInV used as an EIEC diagnostic probe (*ial*).<sup>15</sup>

**Table 2: Virulence loci of 80 *Escherichia coli* O104:H4 outbreak isolates by pathogroup, compared with HUSEC041 and enteroaggregative *E. coli* strain 55989**

Bielaszewska et al., Lancet ID, 2011

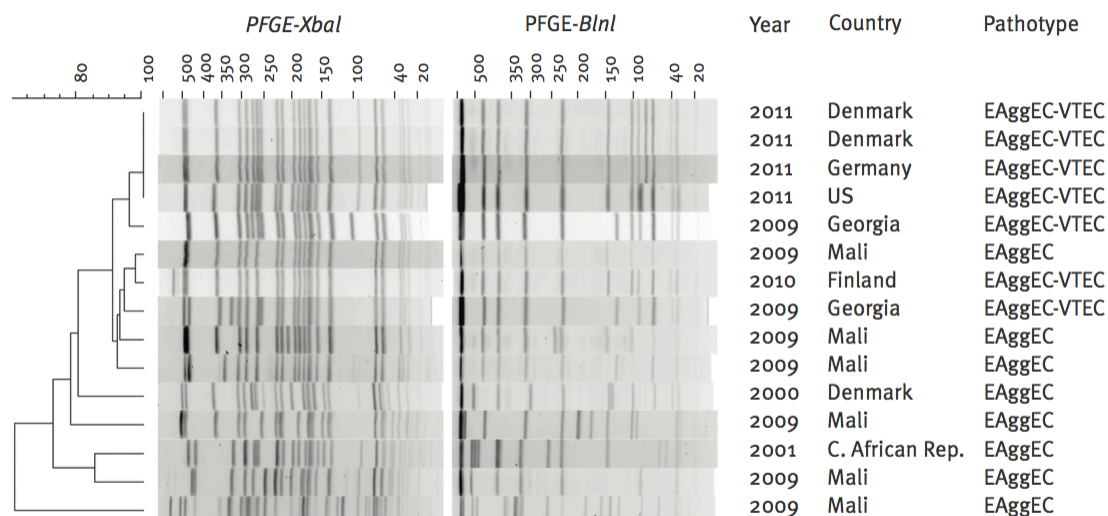
# Outbreak isolates are indistinguishable by conventional molecular epidemiology

Product		Presence in <i>E coli</i> O104:H4*	Presence in HUSEC041	Presence in EAEC 55989†
STEC				
stx <sub>1</sub>	Shiga toxin 1	0/80	No	No
stx <sub>2</sub>	Shiga toxin 2	80/80	Yes	No
EHEC-hlyA	EHEC haemolysin	0/80	No	No
cdt (I-V)	Cytolethal distending toxin	0/80	No	No
subAB	Subtilase cytotoxin	0/80	No	No
espP	Serine protease EspP	0/80	No	No
eae	Intimin	0/80	No	No
iha	Iha (IraA homologue adhesin)	80/80	Yes	Yes

**FIGURE**

PFGE profiles (*Xba*I and *Bln*I) of *Escherichia coli* O104 compared with four isolates from the outbreak of haemolytic uremic syndrome in Germany, May to June 2011

Virulence factors  
MLST  
PFGE



C. African Rep: Central African Republic; CDC: Centers for Disease Control and Prevention; PFGE: pulsed-field gel electrophoresis; RKI: Robert Koch Institute; US: United States.

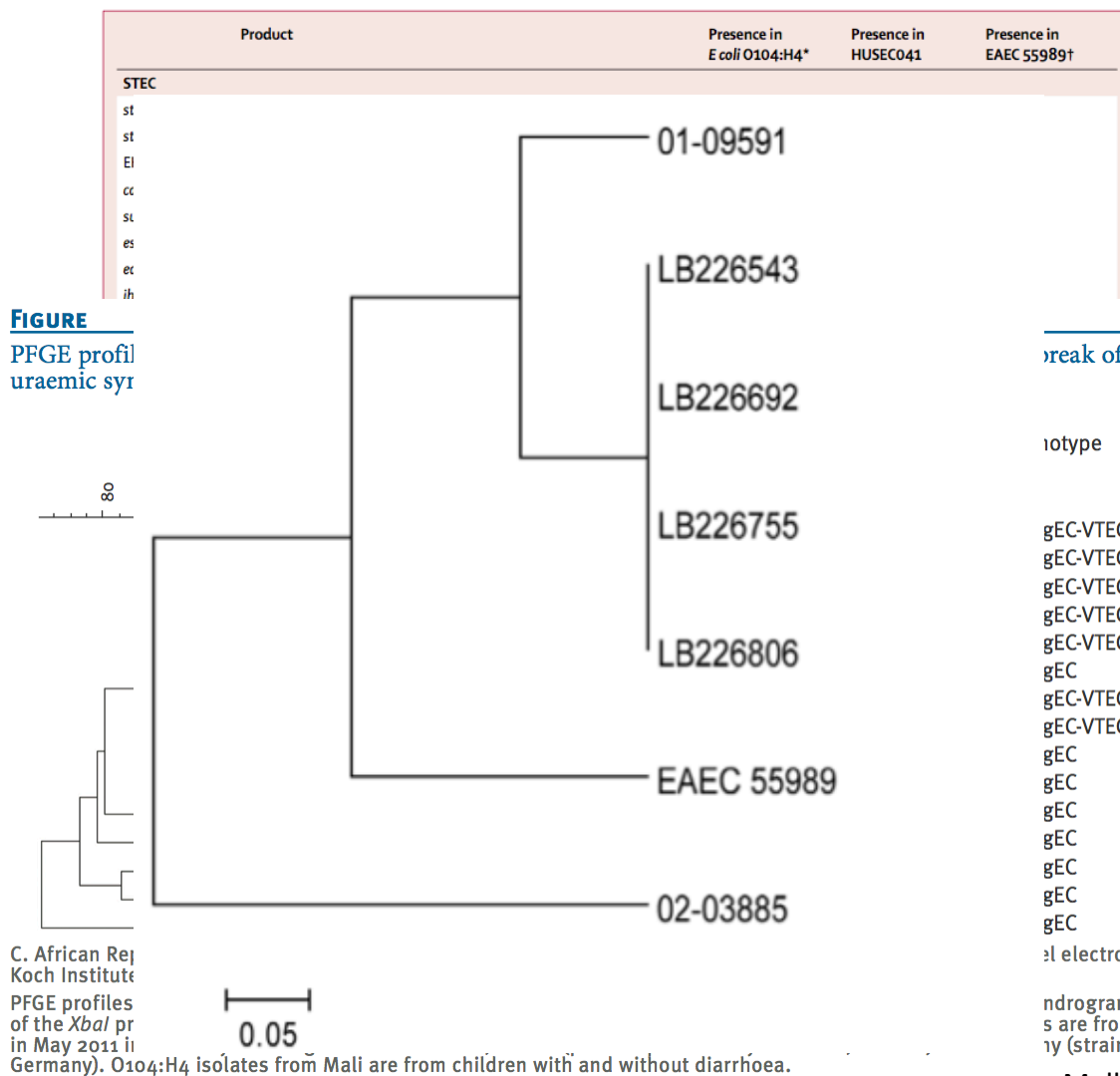
PFGE profiles (*Xba*I and *Bln*I) using the *E. coli* non-O157 PulseNet protocol ([www.pulsenetinternational.com](http://www.pulsenetinternational.com)). Dendrogram based on analysis of the *Xba*I profiles. All isolates are EAggEC O104:H4 with and without *stx*<sub>2</sub>/*vtx*<sub>2</sub> gene. German outbreak isolates are from patients infected in May 2011 in Germany and diagnosed in Denmark, the US (profiles provided by PulseNet, US CDC) and Germany (strain provided by RKI, Germany). O104:H4 isolates from Mali are from children with and without diarrhoea.

Table 2: Virulence loci of 80 *Escherichia coli* O104:H4 outbreak isolates by pathogroup, compared with HUSEC041 and enteroaggregative *E. coli* strain 55989

Scheutz et al., Eurosurveillance, 2011



# Outbreak isolates are indistinguishable by conventional molecular epidemiology



# Virulence factors

## MLST

## PFGE

## Optical mapping

gEC-VTEC  
gEC-VTEC  
gEC-VTEC  
gEC-VTEC  
gEC-VTEC  
gEC  
gEC-VTEC  
gEC-VTEC  
gEC  
gEC  
gEC  
gEC  
gEC  
gEC  
gEC  
gEC

ndrogram based on analysis  
s are from patients infected  
y (strain provided by RKI,

Mellmann et al., PLoS One, 2011

Table 2: Virulence loci of 80 *Escherichia coli* O104:H4 outbreak isolates by pathogroup, compared with HUSEC041 and enteroaggregative *E. coli* strain 55989

# **Outbreak in France...**

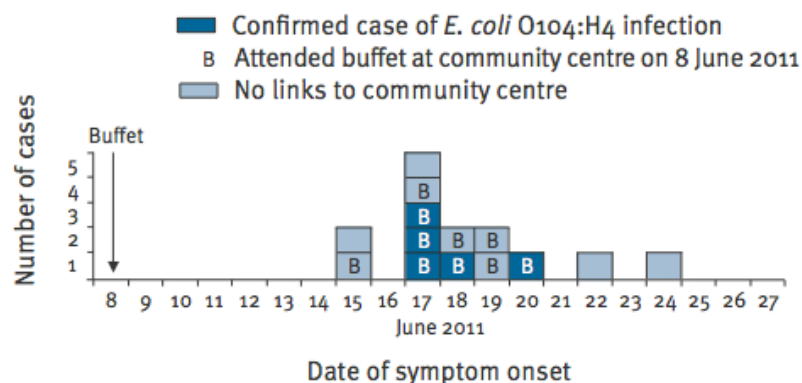
## RAPID COMMUNICATIONS

# Outbreak of haemolytic uraemic syndrome and bloody diarrhoea due to *Escherichia coli* O104:H4, south-west France, June 2011

G Gault<sup>1</sup>, F X Weill<sup>2</sup>, P Mariani-Kurkdjian<sup>3</sup>, N Jourdan-da Silva<sup>4</sup>, L King<sup>4</sup>, B Aldabe<sup>1</sup>, M Charron<sup>1</sup>, N Ong<sup>1</sup>, C Castor<sup>1</sup>, M Macé<sup>2</sup>, E Bingen<sup>3</sup>, H Noël<sup>4</sup>, V Vaillant<sup>4</sup>, A Bone (a.bone@invs.sante.fr)<sup>4,5</sup>, B Vendrely<sup>6</sup>, Y Delmas<sup>6</sup>, C Combe<sup>6</sup>, R Bercion<sup>7</sup>, E d'Andigné<sup>7</sup>, M Desjardin<sup>7</sup>, H de Valk<sup>4</sup>, P Rolland<sup>1</sup>

### FIGURE

Cases of HUS or bloody diarrhoea due to enterohaemorrhagic *Escherichia coli* O104:H4 with date of symptom onset since 10 June 2011, Gironde, France, June 2011 (n=14)



HUS: haemolytic uraemic syndrome.

Of the 15 cases of HUS or bloody diarrhoea, date of symptom onset was unavailable for one case, who attended the buffet on 8 June 2011.

Identical by PFGE, MLST, virulence factors to the German outbreak

# How are these outbreaks related?



## TECHNICAL REPORT OF EFSA

**Tracing seeds, in particular fenugreek (*Trigonella foenum-graecum*) seeds, in relation to the Shiga toxin-producing *E. coli* (STEC) O104:H4 2011 Outbreaks in Germany and France<sup>1</sup>**

**European Food Safety Authority<sup>2,3</sup>**

European Food Safety Authority (EFSA), Parma, Italy

15,000 kg Fenugreek seeds  
Exported from Egypt  
11/24/2009  
Arrived in Germany  
12/14/2009



400 kg  
To a seed supplier /  
repacker in England

10,500 kg  
To a distributor in Germany



50g packets of seeds to a  
distributor in France  
January, 2011

75 kg to sprout farm in  
Lower Saxony  
February, 2011

One cluster in France

41 clusters in Germany



# Approach

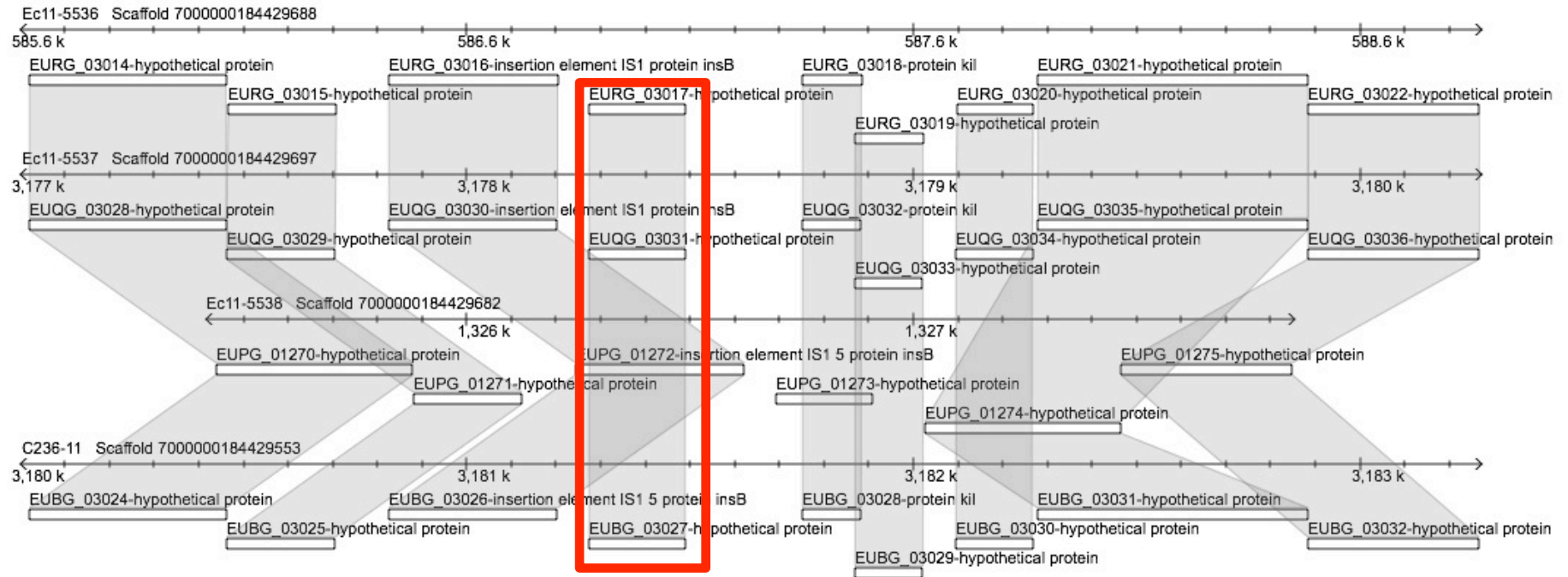
- Multi-platform sequencing
- Assessment of diversity by:
  - SNP calling as compared to TY-2482 (BGI sequenced isolate; Rohde et al., 2011)
  - Variation in gene content



# Data Generated

Isolate Name	Outbreak Location	Syptom Date	Isolation Date	Illumina	454	PacBio
Ec04-8351	-	2004	2004	✓		✓
Ec09-7901	-	2009	2009	✓		✓
Ec11-3677	Germany	May 21, 2011	May 21, 2011	✓		✓
Ec11-3798	Germany	May 21, 2011	May 21, 2011	✓		
C227-11	Germany	Unavailable	May 18, 2011	✓	✓	✓
C236-11	Germany	Unavailable	May 21, 2011	✓	✓	✓
Ec11-4404	France	June 17, 2011	June 21, 2011	✓		
Ec11-5536	France	June 17, 2011	June 24, 2011	✓		
Ec11-5537	France	June 20, 2011	June 24, 2011	✓		
Ec11-5538	France	June 20, 2011	June 24, 2011	✓		
Ec11-4632_C1	France	June 15, 2011	June 25, 2011	✓		
Ec11-4632_C2	France	June 15, 2011	June 25, 2011	✓		
Ec11-4632_C3	France	June 15, 2011	June 25, 2011	✓		
Ec11-4632_C4	France	June 15, 2011	June 25, 2011	✓		
Ec11-4632_C5	France	June 15, 2011	June 25, 2011	✓		
Ec11-4623	France	June 18, 2011	June 27, 2011	✓		
Ec11-4522	France	June 18, 2011	June 22, 2011	✓		

# Single gene



- Plasmid variable cluster not confirmed by PCR
- Chromosomal variable cluster confirmed by PCR

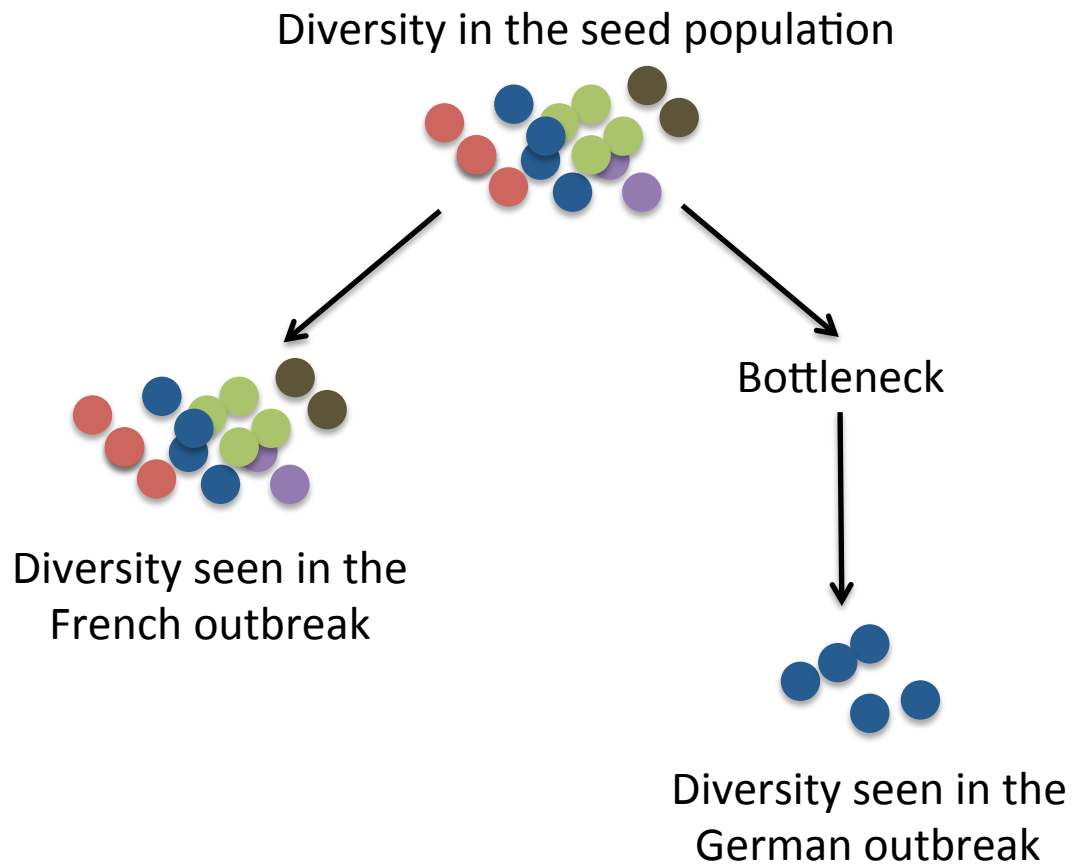
# SNP results summary

- Single SNP observed in 2/4 German isolated strains
- No SNP observed in other two strains
- 19 SNP observed in 11 French isolates and include single German SNP
- Sanger resequencing shows 57 additional German isolates lack diversity observed in Bordeaux samples

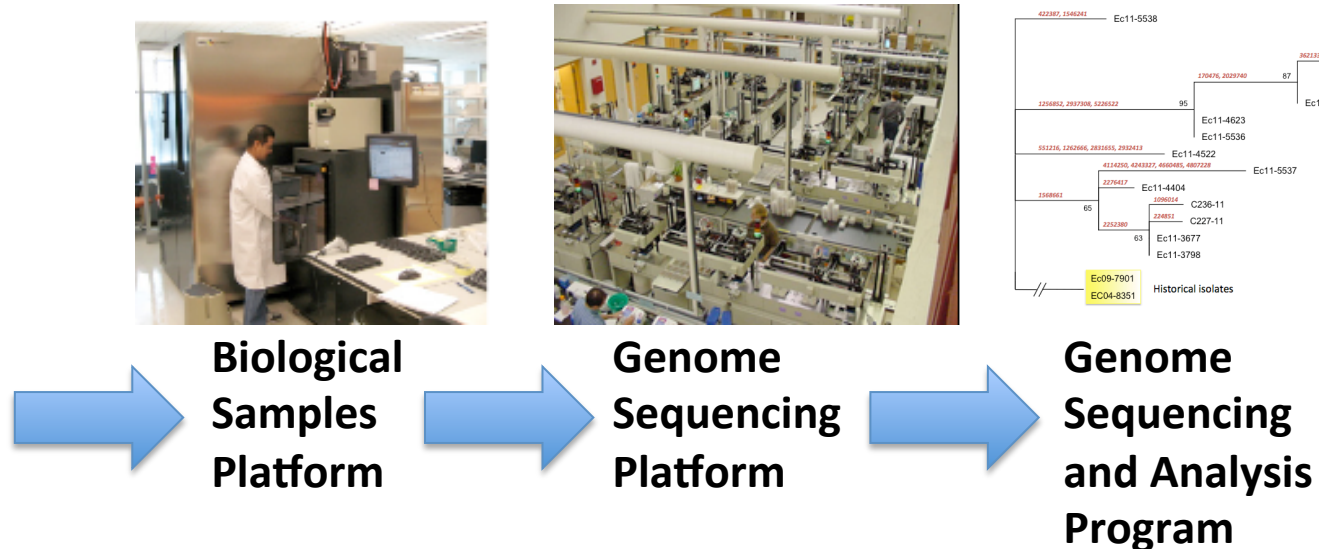
**French outbreak shows increased variation compared with German isolates. Additional WGS pending on larger German sample**



# Hypothesis: Bottleneck in German outbreak



# Standard Broad Path



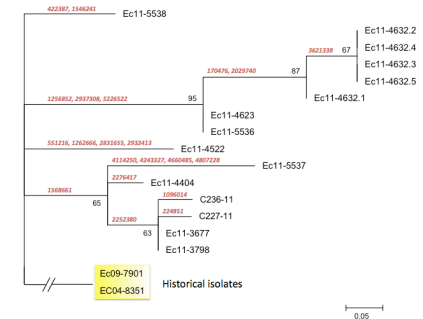
- Meta data entry
- Quantitation
- Sample QC

- Library construction
- Sequencing
- Raw data analysis

- assembly
- SNP calls
- annotation
- release
- analysis

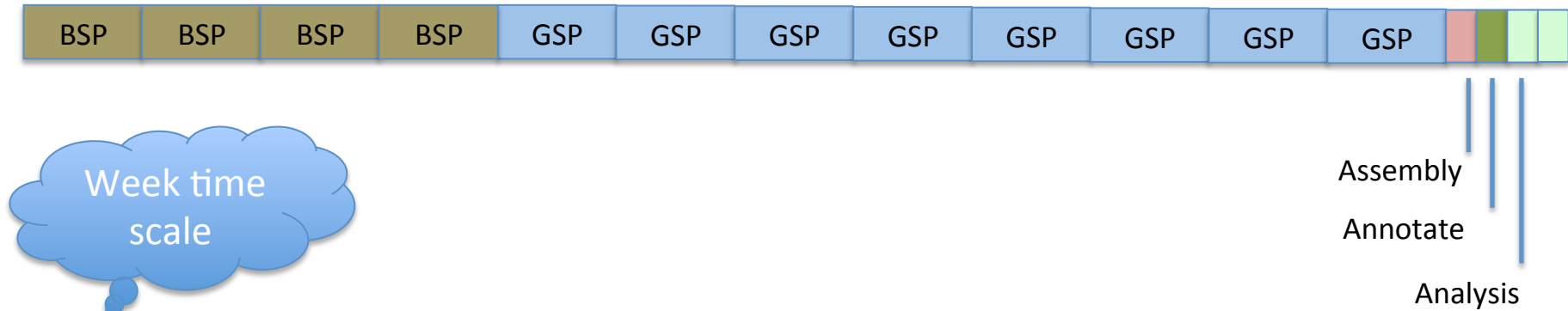
146,966  
samples  
collected

4787  
WGS



# Process time line

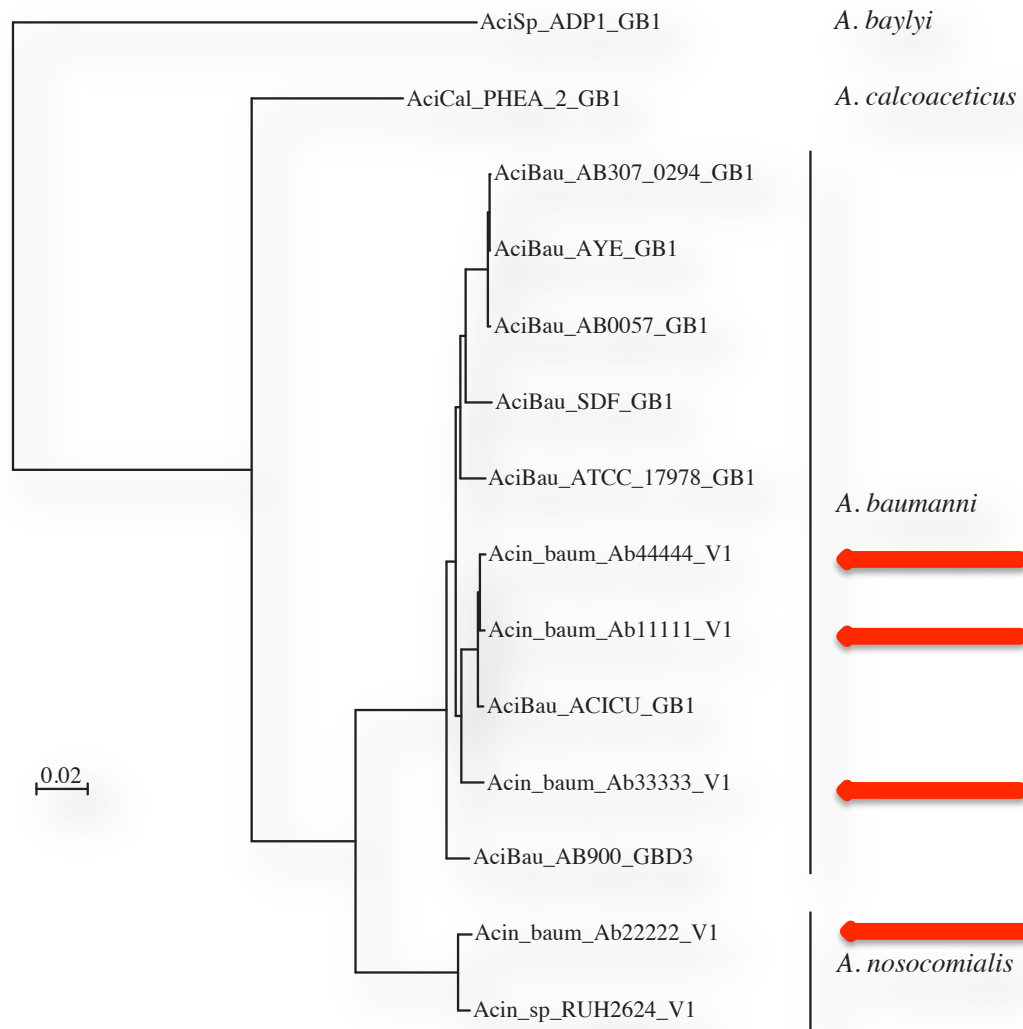
**12 + weeks standard pipeline time**



**11 days for current fire drill pipeline**



# Acinetobacter results



Based on 1900 single copy genes

# Available parts

- Sequencing technology (MiSeq, HiSeq, PacBio)
- Assembly, improvement and SNPs
- Annotation (includes Genbank submission)
- Analysis
  - gene content
  - phylogeny
- Sequence structure
  - Experimental SNP or indel confirmation



# Summary

- We have developed a rapid bacterial sequencing and analysis system that can produce rich data sets in under two weeks.
- It includes numerous off the shelf parts that can be deployed depending on need.
- Whole genome sequence uncovered variation invisible to traditional methods.
- We observed limited SNP variation in German E. coli O104:H4 isolates compared to a smaller French outbreak.
- The system has allowed us to spot issues with isolate nomenclature for burn patient Acinetobacter isolates.

# Acknowledgements

**Genome sequencing  
platform**

**Biological samples  
platform**

## Funding:

This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract Number HHSN272200900018C (BWB), and the Infectious Disease Program of the Broad Institute. ML and WPH received support from Award Number U54GM088558 from the National Institute of General Medical Sciences. YHG received support from National Institutes of Allergy and Infectious Disease (T32 grant AI007061). KAK was partially supported by the Danish Council for Strategic Research (09-063070).

Thank you

EMPLOYEES MUST  
WASH HANDS

